1　前回に引き続き

1.1　正則化付き最適化問題の最適性条件（再掲）

$$\underset{\boldsymbol{w}\in\mathbb{R}^p}{\text{minimize}} \quad f(\boldsymbol{w}) + \lambda\Omega(\boldsymbol{w}) \tag{1.1}$$

に対し, あるベクトル $\boldsymbol{w}\in\mathbb{R}^p$ が最適解となる必要十分条件は以下の式を満たす $-\frac{1}{\lambda}\nabla f(\boldsymbol{w}) \in \partial\Omega(\boldsymbol{w})$ である.

$$\partial\Omega(\boldsymbol{w}) = \begin{cases} \{\boldsymbol{z}\in\mathbb{R}^p \mid \Omega^*(\boldsymbol{z}) \le 1\} & \text{if } \boldsymbol{w}=0 \\ \{\boldsymbol{z}\in\mathbb{R}^p \mid \Omega^*(\boldsymbol{z}) = 1 \text{ and } \boldsymbol{z}^\top\boldsymbol{w} = \Omega(\boldsymbol{w})\} & \text{otherwise.} \end{cases} \tag{1.2}$$

次の Fenchel-Young Inequality を用いて式（??）を証明できる.

Fenchel-Young Inequality

定理 1.1. $\boldsymbol{w}\in\mathbb{R}^p, f\in\mathbb{R}^p, \boldsymbol{z}\in\mathbf{dom}\, f^* \neq \emptyset$ とおくと，次の不等式

$$f(\boldsymbol{w}) + f^*(\boldsymbol{z}) \ge \boldsymbol{w}^\top\boldsymbol{z}$$

が成り立ち，また，$\boldsymbol{z}\in\partial f(\boldsymbol{w})$ の時等号が成立.

証明は板書で.

1.2　ノルムに対しての Fenchel 共役の双対性

前回引き続いて，次の不等式の双対性を示す.

$$\max_{\boldsymbol{z}\in\mathbb{R}^p, \Omega^*(\boldsymbol{z})\le\lambda} -f^*(\boldsymbol{z}) \le \min_{\boldsymbol{w}\in\mathbb{R}^p} f(\boldsymbol{w}) + \lambda\Omega(\boldsymbol{w}) \tag{1.3}$$

関数 f の定義域が空でないとき等号が成り立つ.
弱双対性は Fenchel-Young Inequality より証明され，強双対性は本 [?] の定理 3.3.5 を参照．板書.

1.3　双対性を用いて問題（??）を解く

$z(\boldsymbol{w}^*) = \nabla f(\boldsymbol{w}^*)$ は双対問題の唯一の解である．一般的には，$\boldsymbol{z} = \min(1, \frac{1}{\Omega^*(\nabla f(\boldsymbol{w}))})\nabla f(\boldsymbol{w})$ のような $\boldsymbol{z}$ を取り，各反復で $\boldsymbol{z}(\boldsymbol{w}^*)$ を最適解して双対ギャップをゼロにする.

1.4　Design Matrix を用いた定式化

関数 f を関数 $\psi:\mathbb{R}^n\to\mathbb{R}$，Design Matrix $\boldsymbol{X}$ を用いて $f(\boldsymbol{w}) = \phi(\boldsymbol{X}\boldsymbol{w})$ と変換する.

$$\min_{\boldsymbol{w}\in\mathbb{R}^p, \boldsymbol{u}\in\mathbb{R}^n} \quad \phi(\boldsymbol{u}) + \lambda\Omega(\boldsymbol{w}) \tag{1.4}$$

$$\text{subject to} \quad \boldsymbol{u} = \boldsymbol{X}\boldsymbol{w} \tag{1.5}$$

また，この問題をラグランジュ乗数 $\boldsymbol{\alpha}^\intercal \in \mathbb{R}^n$ を用いてラグランジュ双対問題に変換すると

$$\min_{\boldsymbol{w}\in\mathbb{R}^p, \boldsymbol{u}\in\mathbb{R}^n} \max_{\boldsymbol{\alpha}^\intercal\in\mathbb{R}^n} \phi(\boldsymbol{u}) + \lambda\Omega(\boldsymbol{w}) + \lambda\boldsymbol{\alpha}^\intercal(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{u}) \tag{1.6}$$

$$\min_{\boldsymbol{w}\in\mathbb{R}^p, \boldsymbol{u}\in\mathbb{R}^n} \max_{\boldsymbol{\alpha}^\intercal\in\mathbb{R}^n} (\phi(\boldsymbol{u}) - \lambda\boldsymbol{\alpha}^\intercal\boldsymbol{u}) + \lambda(\Omega(\boldsymbol{w}) + \lambda\boldsymbol{\alpha}^\intercal\boldsymbol{X}\boldsymbol{w}) \tag{1.7}$$

そして，Fenchel 双対問題は

$$\max_{\boldsymbol{\alpha}\in\mathbb{R}^n} -\phi^*(\lambda\boldsymbol{\alpha}) \ \text{ such that } \ \Omega^*(\boldsymbol{X}^\intercal\boldsymbol{\alpha}) \leq \lambda \tag{1.8}$$

2 汎用メソッド

$$\underset{\boldsymbol{w}\in\mathbb{R}^p}{\text{minimize}} \quad f(\boldsymbol{w}) + \lambda\Omega(\boldsymbol{w}) \tag{2.1}$$

損失関数 f と正則化項 Ω が同時に凸である場合だけ式（**??**）は凸である．

2.1 劣勾配法

もし問題の劣勾配を効率よく算出できれば，すべての制約なし凸問題が劣勾配法で解ける．式 (**??**) に対して，損失関数 f と正則化項 Ω の劣勾配が計算できれば劣勾配法で解ける．

その反復アルゴリズムは：

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \frac{\alpha}{t}(\boldsymbol{s} + \lambda\boldsymbol{s}'), \quad \text{where } \boldsymbol{s} \in \partial f(\boldsymbol{w}_t), \boldsymbol{s}' \in \partial\Omega(\boldsymbol{w}_t) \tag{2.2}$$

本 [**?**] より，この反復アルゴリズムが大域的収束，収束率は $F(\boldsymbol{w}_t) - \min_{\boldsymbol{w}\in\mathbb{R}^p} F(\boldsymbol{w}) = O(\frac{1}{\sqrt{t}})$，だが，実際に劣勾配法の収束スピードが遅いにもかかわらず，スパース解を求めるのも難しい．

2.2 LP,QP,SOCP,SDP 問題に定着

この一章すべての正則化付き最適化問題（正則化付き最小二乗問題）は SDP 問題または QP など SDP より簡単な問題に定着できる．

例えば，次の問題

$$\min_{\boldsymbol{w}\in\mathbb{R}^p} \frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 + \lambda\Omega(\boldsymbol{w}) \tag{2.3}$$

の実数制約を非負制約に変形すると，次の QP 問題に定着できる

$$\min_{\boldsymbol{w}_+, \boldsymbol{w}_-\in\mathbb{R}^p} \frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}_+ + \boldsymbol{X}\boldsymbol{w}_-\|_2^2 + \lambda\Omega(1^\intercal\boldsymbol{w}_+ + 1^\intercal\boldsymbol{w}_-) \tag{2.4}$$

汎用ツールを使用すれば高精度解（小さい双対ギャップ）を求められるが，機械学習にとっては効率悪くて贅沢すぎる．なぜなら，

1. これらのツールが問題の特徴を無視しスピードが遅い，メモリー不足で落ちる可能性もある．
2. Bottou and Bousquet(2007) より，機械学習に対して高精度の解がいらない．

参考文献

[1] Francis Bach, Rodolphe Jenatton, EJulien Mairal and Guillaume Obozinski (2012) "Optimization with Sparsity-Inducing Penalties", Foundations and Trends in Machine Learning: Vol. 4: No 1, pp 1-106.

[2] 福島雅夫, 非線形最適化の基礎, 朝倉書店, 2001

[3] Borwein, Jonathan; Lewis, Adrian (2006). Convex Analysis and Nonlinear Optimization: Theory and Examples (2 ed.). Springer. ISBN 978-0-387-29570-1.

[4] Y. Nesterov, Introductory Lectures on Convex Optimization: A Basic Course (Applied Optimization), 1st ed. Springer Netherlands.

[5] Steve Wright, NIPS Tutorial, 6 December 2010, `http://pages.cs.wisc.edu/~swright/nips2010/sjw-nips10.pdf`